Time-Stampless Adaptive Nonuniform Sampling for Stochastic Signals

Soheil Feizi, Vivek K Goyal, *Senior Member, IEEE*, and Muriel Médard, *Fellow, IEEE*

***Abstract*—In this paper, we introduce a *time-stampless adaptive nonuniform sampling* (TANS) framework, in which time increments between samples are determined by a function of the m most recent increments and sample values. Since only past samples are used in computing time increments, it is not necessary to save sampling times (time stamps) for use in the reconstruction process. We focus on two TANS schemes for discrete-time stochastic signals: a greedy method, and a method based on dynamic programming. We analyze the performances of these schemes by computing (or bounding) their trade-offs between sampling rate and expected reconstruction distortion for autoregressive and Markovian signals. Simulation results support the analysis of the sampling schemes. We show that by opportunistically adapting to local signal characteristics TANS may lead to improved power efficiency in some applications.**

I. Introduction

Sampling is essential in any digital system that interfaces with the analog world. All else being equal, it is desirable to minimize the number of samples while maintaining an acceptable reconstruction distortion. In some applications, minimizing the number of samples can be translated into having a power-efficient sampling, since the power consumption at an analog-to-digital converter (ADC) is approximately proportional to its sampling rate [1]. Also, having fewer samples can increase the efficiency of other processing of these measurements. For example, if these samples should be transmitted to another place via a communication channel, having fewer samples will improve power and bandwidth efficiencies.

One can view *sampling* as a *query* to obtain information from a signal or a function that can only be measured remotely. We get a sample of this function at an arbitrary time when we *query* for it over a communication medium. Here, a portion of the operational cost (e.g., power) is proportional to the number of samples that we acquire. Hence, again it is desirable to minimize the number of samples taken.

A uniform sampling at the Nyquist rate of the signal may cause some redundant samples, since the global signal bandwidth may not be a good measure of local variations of the signal. Although traditional nonuniform sampling schemes (e.g., [2], [3]) deal with this problem, they have certain limitations. Firstly, they are mostly designed to operate under specific conditions for restrictive signal models (e.g., [4]–[8]) and, secondly, sampling times (i.e., time stamps) must be stored/transmitted to be used in the reconstruction process. This may cause power/bandwidth inefficiencies in sampling/communication procedures.

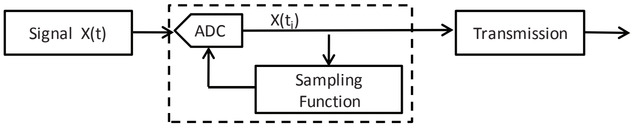

Fig. 1. A schematic view of the TANS framework: sampling times are determined by a function of m most recently taken samples. Hence, it is not necessary to save sampling times (time stamps) for use in the reconstruction process.

In this paper, we introduce a new framework for an adaptive nonuniform sampling scheme (see Figure 1). The key idea of this framework is that *time increments between samples are computed by using a function of previously taken samples*. Therefore, keeping sampling times (time stamps), except initialization times, is not necessary. The function by which sampling time intervals is computed is called the *sampling function*. The aim of this sampling framework is to have a balance between the reconstruction distortion and the average sampling rate. We refer to this sampling framework as *Time-stampless Adaptive Nonuniform Sampling* (TANS). The TANS concept can be applied on continuous- or discrete-time signals, and the design and analysis can be based on deterministic or stochastic models.

The TANS framework is described in general terms in Section II. Section III then formalizes the problem setup for this paper, where we focus on discrete-time signals and stochastic models. A greedy method is developed in Section IV, and a method based on dynamic programming (DP) is developed in Section V. Simulations results are provided in Section VI, and some proofs are relegated to Section VII. Section VIII concludes the paper.

II. TANS Framework

In this section, we introduce the TANS framework and sketch design approaches that will be developed in more detail in later sections. Fix some nonnegative integer m and suppose the ith sample of signal $X(t)$ is taken at time t_i. We take the $(i+1)$st sample after a time increment of

$$T_i = f\left(\{(t_j, X(t_j)) : i-m+1 \le j \le i\}\right),$$

This material is based upon work supported in part by the National Science Foundation under Grant No. 0643836, the Texas Instruments Leadership University Program and AFOSR under award No. 016974-002. The material in this paper was present in part at 2010 Annual Allerton Conference on Communication, Control, and Computing.

The authors are with the Research Laboratory of Electronics and the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (email: {sfeizi,vgoyal,medard}@mit.edu). Telephone: +1 617 253 3167.

where f is called the sampling function. This makes the sampling rate adapt to local characteristics of the signal. Since the time increment is a function of the m most recently taken samples, we say the *order* of the sampling function f is m. The sampling is nonuniform except in the trivial cases when f is a constant-valued function (e.g., $m = 0$). Some initialization of the first m sampling times is necessary, but the effect of this initialization on the rate is amortized.

The sampling function is known at the reconstruction side. Assuming that the *state*

$$S_{t_i} = \{(t_j, X(t_j)) : i - m + 1 \leq j \leq i\}$$

is also known at the reconstruction side when reconstructing $X(t)$ on $[t_i, t_{i+1}]$, there is *no need for the sampling times (time stamps) to be transmitted*. These times can be computed by using the sampling function and previously taken samples:

$$t_{i+1} = t_i + f\left(S_{t_i}\right).$$

This type of synchronization in an adaptive system without explicit communication is often called backward adaptation [9]. In a practical setting involving both sampling and quantization, backward adaptivity requires using the quantized values to drive the adaptation [10]–[13]. Here, to maintain focus on sampling rate and adaptation of sampling increments, we do not explicitly include quantization effects. Note that while the sampling time selection is causal, the reconstruction method can be causal or non-causal.

The aim of TANS is to balance between the average sampling rate and the reconstruction distortion. This objective is different from the one considered in change point analysis [14] or active learning [15]. There, the objective is to find points of the signal at which statistical behaviors of the signal change, by causal or non-causal sampling, respectively.

Suppose $\widehat{X}(t)$ is the reconstructed signal computed by some reconstruction method. For the case of discrete time and a stochastic signal model, define $d(S_{t_i}, T_i)$ as the expected reconstruction distortion over samples from time $t_i + 1$ until time $t_{i+1} - 1$. That is,

$$d(S_{t_i}, T_i) = \mathbb{E}_{\mathcal{X}}\left[\sum_{t=t_i+1}^{t_{i+1}-1} D\big(X(t), \widehat{X}(t)\big)\right],$$

where $\mathcal{X}$ is the known probabilistic model of the signal $X(t)$ and $D(X(t), \widehat{X}(t))$ represents the distortion at time t.[1] Note that at times t_i and t_{i+1} the reconstruction distortion is zero since exact sample values are known at these times. In realistic cases and for a given state S_{t_i}, $d(S_{t_i}, T_i)$ is an increasing function with respect to T_i, because the greater the next sampling step, the greater the reconstruction distortion. On the other hand, the greater the next sampling step, the larger the rate benefit. Hence, a rate penalty can be defined as $a(S_{t_i}, T_i) = \rho/f(S_{t_i}) = \rho/T_i$, where ρ is a rate award parameter. We define the cost of each sampling state as the sum of the expected reconstruction distortion and the rate penalty, that is, $c(S_{t_i}, T_i) = d(S_{t_i}, T_i) + a(S_{t_i}, T_i)$. The overall cost of the sampling process is the sum of different sampling state costs, that is, $\sum_i c(S_{t_i}, T_i)$.

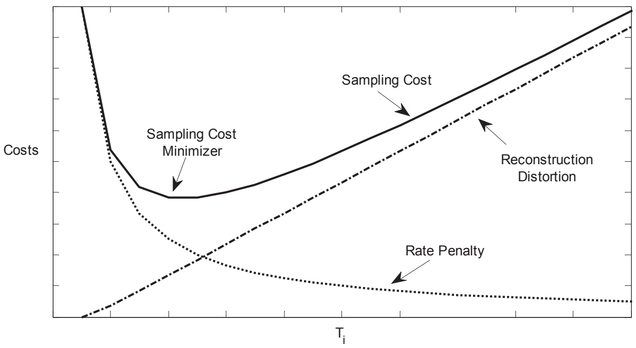

Fig. 2. Demonstrating the behavior of different parts of sampling state cost $c(S_{t_i}, T_i) = d(S_{t_i}, T_i) + a(S_{t_i}, T_i)$, where $d(S_{t_i}, T_i)$ is the reconstruction distortion and $a(S_{t_i}, T_i)$ is the rate penalty function.

Finding an appropriate sampling function for TANS depends on requirements such as average sampling rate, maximum distortion, etc. In this paper, we investigate two general approaches to computing appropriate sampling functions for given sampling setups: *greedy methods* and *dynamic programming (DP) methods*.

In greedy methods, a sampling function at state S_{t_i} chooses the next sampling increment T_i to minimize the sampling state cost $c(S_{t_i}, T_i)$. As depicted in Figure 2, for any given state S_{t_i}, $d(S_{t_i}, T_i)$ is an increasing function of T_i, while $a(S_{t_i}, T_i)$ is a decreasing function. Therefore, there is a trade-off between the sampling rate and the expected reconstruction distortion. A greedy method balances this trade-off by choosing the state cost minimizer as the next sampling increment. In certain cases, greedy sampling schemes can perform closely to an optimal scheme.

Since $c(S_{t_i}, T_i)$ depends on the current sampling state S_{t_i}, a greedy sampling function does not take into account characteristics of the next sampling state. Intuitively, the larger the sampling increment T_i at the sampling state S_{t_i}, the lower the *quality* of the next sampling state. Hence, in general, greedy methods are not optimal sampling schemes considering the overall sampling cost $\sum_i c(S_{t_i}, T_i)$ as a comparison measure. We consider effects of the next sampling states' *quality* in DP methods. We show that an exact Bellman-Ford equation (BFE) can be written and solved for some sampling setups. For those cases, the solution of BFE provides an optimal sampling function which minimizes the overall sampling cost. In cases where solving the BFE is not practically feasible (because the number of possible sampling states is large or exact sampling states are not known), we propose sampling functions based on approximate dynamic programming (ADP) algorithms. Sampling functions derived by greedy methods can be used in ADP-based sampling methods. In fact, greedy methods can be viewed as DP-based methods with a unit time horizon.

In this paper, we consider two examples of stochastic signals: stationary signals and Markovian signals. Unlike stationary signals, Markovian processes have sudden changes

[1] An analogous formulation for continuous time would replace the sum with an integral over $t \in [t_i, t_{i+1}]$. Without a stochastic model, a maximum error criterion could be used.

in their statistical properties based on an underlying hidden Markov chain. Note that the general sampling framework can be applied on other signal models.

III. Problem Setup, Signal Models and Background Results

In this section, we first present the problem setup of TANS. Then, we introduce signal models considered in this paper. At the end of this section, a *generalized linear prediction filter* is proposed, which linearly predicts the future samples of a stationary process by using a set of nonuniform samples from its past.

A. Problem Setup

Consider a discrete-time signal $X(t)$. TANS with order m is used to take samples from this signal. Hence, the next sampling increment at time t_i is a function of m most recently taken samples at that time (i.e., $T_i = f(S_{t_i})$). The set of samples taken using the sampling function $f(\cdot)$ is denoted by $\Gamma^f_{X(t)}$. A function $g(\cdot)$ is used to reconstruct the original signal from its samples, that is, $\widehat{X}(t) = g(\Gamma^f_{X(t)})$. The reconstruction error signal is $e(t) = X(t) - \widehat{X}(t)$. The overall sampling cost is the sum of sampling state costs over different states, that is, $c_{\rm tot}(X(t), f, g) = \sum_i c(S_{t_i}, T_i)$.

A system optimization problem under the TANS framework can be stated as follows:

Definition 1 (Optimal TANS sampling problem): For a class of signals $X(t)$, a given reconstruction function $g(\cdot)$, and an order m, a sampling function $f^*(\cdot)$ is desired to minimize the overall expected sampling cost:

$$f^* = \arg\inf_f c_{\rm tot}(f, g). \tag{1}$$

The resulting cost is denoted $c^*_{\rm tot} = c_{\rm tot}(f^*, g)$.

B. Signal Models

In this paper, we consider the following signal models:

- Case 1: an autoregressive signal with memory of one (i.e., AR(1)):

$$X(t+1) = \alpha X(t) + Z(t+1), \tag{2}$$

where $Z(t+1)$ is a Gaussian noise with zero mean. If the power of the signal is assumed to be one, the noise variance is $1-\alpha^2$.

- Case 2: a Markovian signal:

$$X(t+1) = \alpha_{\theta_t} X(t) + Z_{\theta_t}(t+1), \tag{3}$$

where θ_t represents the state of a hidden Markov chain (MC) with state transition probabilities depicted in Figure 3. At time t, if the MC is at state 0, $\theta_t = 0$; otherwise, $\theta_t = 1$. Depending on the value of θ_t, the signal is generated by a first-order AR model with parameter α_{θ_t} and the noise variance $1-\alpha^2_{\theta_t}$. Note that, in this model, unlike the previous case, the coefficient of the AR model has a sudden change in time depending on the state of the underlying hidden Markov chain.

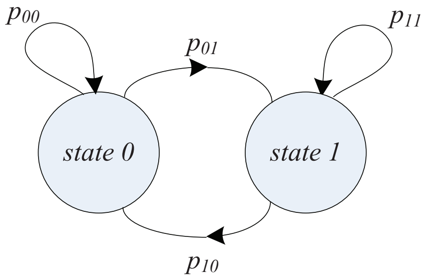

Fig. 3. A hidden Markov chain considered in Markovian signal model of equation (3).

C. Generalized Linear Prediction Filter

Suppose $X(t)$ is a stationary signal. Assume we have m samples of $X(t)$ at times $\{t_{i-m+1}, \ldots, t_i\}$. Our aim is to linearly predict $X(t)$, where $t = t_i + T$ for some $T \geq 1$, by using these known samples so that the expected mean square error is minimized (MMSE predictor).

Consider $\widehat{X}(t)$ as a predicted value of $X(t)$ by using these m sample values. The prediction error is $e(t) = X(t) - \widehat{X}(t)$. Define $\tau_k = t - t_{i-k}$ for $0 \leq k \leq m-1$. We want to find optimal linear prediction weights $w_{\tau_0}, \ldots, w_{\tau_{m-1}}$ so that the prediction error power is minimized:

$$\begin{aligned} \min_{w_{\tau_0},\ldots,w_{\tau_{m-1}}} \quad & E[|e(t)|^2] \\ \text{subject to} \quad & \widehat{X}(t) = \sum_{k=0}^{m-1} w_{\tau_k} X(t_{i-k}). \end{aligned} \tag{4}$$

A solution of this linear optimization is referred as $w^*_{\tau_i}$, for $0 \leq i \leq m-1$. Note that, unlike a regular linear prediction filter ([16]), a generalized linear prediction filter predicts $X(t)$ by using a set of nonuniform samples.

The auto-correlation function of $X(t)$ can be written as

$$r(i) = E\big[X(t)X^c(t-i)\big], \tag{5}$$

where $X^c(t)$ represents the complex conjugate of $X(t)$. In this paper, we deal with real signals.

To simplify notations, we define the following matrices:

$$\begin{aligned} \tau &= [\tau_0, \ldots, \tau_{m-1}]^T \\ \mathbf{p} &= \big[r(-\tau_0), \ldots, r(-\tau_{m-1})\big]^T \\ \mathbf{w}^*_\tau &= \big[w^*_{\tau_0}, ..., w^*_{\tau_{m-1}}\big]^T. \end{aligned} \tag{6}$$

Also, a $m \times m$ auto-correlation matrix $\mathbf{R}$ is defined whose component in the ith row and jth column is $r(\tau_i - \tau_j)$.

The following theorem provides optimal weights for the generalized linear prediction filter:

Theorem 2:

$$\mathbf{p} = \mathbf{R}\mathbf{w}^*_\tau.$$

Proof: See Section VII-A. ■

For $X(t)$ with zero mean, the variance of the prediction error is defined as follows:

$$\sigma^2_{e^*}(S_{t_i}, T) = E[|e^*(t)|^2].$$

The following theorem provides a way to compute the variance of the prediction error for a generalized linear prediction filter:

Theorem 3: $\sigma^2_{e^*}(S_{t_i}, T) = r(0) - \mathbf{p}^T \mathbf{w}^*_\tau$.

Proof: See Section VII-B. ■

IV. Greedy TANS

In this section, we investigate greedy sampling functions various signal models. In Section V, we evaluate performance of these schemes and compare them with some other schemes, including uniform sampling setups. In all of these sampling schemes, the reconstruction function is assumed to be a generalized linear prediction filter, introduced in Section III-C. Note that it is a causal reconstruction function.

In greedy methods, a sampling function is computed as follows:

$$\begin{aligned} T_i = \arg\min_{T} \quad & c(S_{t_i}, T) \\ \text{subject to} \quad & T \geq 1, \end{aligned} \tag{7}$$

where $f(S_{t_i}) = T_i$. The sampling function of a greedy method depends on the current sampling state and does not take into account characteristics of the next sampling states. Intuitively, the larger the sampling increment T_i at the sampling state S_{t_i}, the lower the *quality* of the next sampling state. Therefore, we have a trade-off between the sampling rate award of the current state and the sampling cost of the next state (Figure 2). Hence, greedy methods usually are not optimal solutions of a TANS sampling problem presented in Definition 1. However, these greedy solutions can be used to approximate optimal solutions, which may have high computational complexity.

Here, we investigate the greedy sampling function of (7) for two signal classes described in Section III-B. We analyze their sampling rate versus the expected reconstruction distortion in Theorems 4 and 8. Simulation results of proposed schemes are shown in Section VI.

A. Greedy TANS for Autoregressive Signals

In this section, we consider greedy sampling functions for an AR(1) signal model described in (2).

Suppose $f^*_{\text{greedy}}(S_{t_i})$ is an optimal greedy sampling function. Also, suppose d^*_{tot} is the expected reconstruction distortion per sample corresponding to this sampling function. The following theorem introduces an optimal greedy sampling function and its expected reconstruction distortion per sample.

Theorem 4: For an AR(1) signal with parameter α described in (2), over a large enough time interval $[0, T_{\text{tot}}]$, an optimal greedy sampling function is

$$\begin{aligned} f^*_{\text{greedy}}(S_{t_i}) &= T^*, \\ d^*_{\text{tot}} &= \frac{1}{T^*} \sum_{j=1}^{T^*-1} 1 - \alpha^{2j}, \end{aligned} \tag{8}$$

where $T^* = \arg\min_T \sum_{j=1}^{T-1}(1-\alpha^{2j}) + \rho/T$.

Proof: See Section VII-C. ■

Note that, for this signal model, an optimal greedy sampling function yields uniform sampling with the sampling rate $1/T^*$, which performs closely to an optimal sampling scheme. In the following corollary, we present a formula to compute T^*.

Corollary 5: Suppose T_{root} is a solution of the following equation:

$$(1-\alpha^{2T}) - \frac{\rho}{T(T+1)} = 0 \tag{9}$$

where $0 < \alpha < 1$ and $\rho > 0$. If $(1-\alpha^2) < \frac{\rho}{2}$, then

(i) T_{root} is unique; and
(ii) $T^* = \lfloor T_{\text{root}} \rfloor$ or $T^* = \lfloor T_{\text{root}} \rfloor + 1$.

Proof: See Section VII-D. ■

We will validate this by simulation in Section VI. Also, arguments of Theorem 4 can be extended for a general stationary signal.

For an AR signal, since statistical properties of the signal do not vary in time, there is no rate adaption with respect to sample values. This is not the case for Markovian signals. We investigate the greedy TANS framework for Markovian signals in the next section; a rate adaption with respect to sample values would be helpful to minimize the sampling cost and leads to a nonuniform sampling scheme.

B. Greedy TANS for Markovian Signals

Consider a Markovian signal described by (3), where θ_t represents the state of a hidden underlying Markov chain depicted in Figure 3. In this section, for simplicity we assume the MC is symmetric (i.e., $p_{01} = p_{10}$). However, all arguments can be extended for a general MC. We also assume that α_0 and α_1 are known. However, the state of the Markov chain (i.e., θ_t) is unknown and needed to be estimated by using the taken samples. We use a generalized linear prediction filter for the reconstruction. Note that if α_0 and α_1 are also unknowns, one can learn these parameters at the beginning of the process by taking more samples. Then, our proposed scheme can be applied for the rest of the process.

Extending the previous notation, define $\theta_{S_{t_i}}$ as the state of the MC during the sampling state S_{t_i}. If during S_{t_i} the MC state stays at zero, $\theta_{S_{t_i}} = 0$. Similarly, if the MC state stays at one, $\theta_{S_{t_i}} = 1$. Otherwise, if there is an MC transition within this sampling state, $\theta_{S_{t_i}} = 2$. We assume that $\theta_{S_{t_i}}$ is unknown and needs to be estimated by using the taken samples. The estimated value of $\theta_{S_{t_i}}$ is referred by $\hat{\theta}_{S_{t_i}}$. The error probability of this estimation is referred by $P_e(S_{t_i}) = Pr(\hat{\theta}_{S_{t_i}} \neq \theta_{S_{t_i}})$.

Algorithm 6: A greedy sampling function for the considered Markovian signal has the following steps:

- Step i,0: Compute $\hat{\theta}_{S_{t_i}}$ and $P_e(S_{t_i})$.
- Step i,1: Compute $T_i = \arg\min_T c(S_{t_i}, T|\hat{\theta}_{S_{t_i}})$, where $c(S_{t_i}, T|\hat{\theta}_{S_{t_i}})$ is the sampling state cost given $\hat{\theta}_{S_{t_i}}$ (see (11) and (12)).
- Step i,2: Take a sample at time $t_i + T_i$.
- Step i,3: Compute S_{t_i+1}. Repeat.

For simplicity, we assume that the sampling increment T is small enough that the probability of having more than one MC transition is negligible. In other words, we have the following assumptions.

Assumption 7:

$$\begin{aligned}\max_i T_i &\leq T_{up},\\ p_{00}^{(T_{up}-1)} &\gg \frac{1}{2},\\ p_{11}^{(T_{up}-1)} &\gg \frac{1}{2}.\end{aligned}$$

If MC transition probabilities p_{01} and p_{10} are small enough, this assumption is reasonable. Under this assumption, the sampling state cost, $c(S_{t_i}, T)$, can be conditioned on the value of $\hat{\theta}_{S_{t_i}}$ as follows:

$$\begin{aligned}&c(S_{t_i}, T\,|\,\hat{\theta}_{S_{t_i}} = 0)\\ &= Pr(\hat{\theta}_{S_{t_i}} = \theta_{S_{t_i}})\Big\{\sum_{j=0}^{T-1} p_{00}^j p_{01} \sum_{l=1}^{j}(1-\alpha_0^{2l})\\ &\quad +(T-1-j)\sigma_{\max}^2\Big\}\\ &\quad +Pr(\hat{\theta}_{S_{t_i}} \neq \theta_{S_{t_i}})\sigma_{\max}^2 + \frac{\rho}{T}\\ &= (1-P_e(S_{t_i}))\Big\{\sum_{j=0}^{T-1} p_{00}^j p_{01} \sum_{l=1}^{j}(1-\alpha_0^{2l})\\ &\quad +(T-1-j)\sigma_{\max}^2\Big\}\\ &\quad +P_e(S_{t_i})(T-1)\sigma_{\max}^2 + \frac{\rho}{T}, \qquad (10)\end{aligned}$$

which says that, if the estimation is correct (with probability $1-P_e(S_{t_i})$) and an MC transition happens at time t_i+j+1 (with probability $p_{00}^j p_{01}$), the sampling state cost is $\sum_{l=1}^{j} 1-\alpha_0^{2l}+(T-1-j)\sigma_{\max}^2$, where $\sigma_{\max}^2$ is the maximum prediction error variance (in this example, $\sigma_{\max}^2 = 1$). If the estimation process fails (with probability $P_e(S_{t_i})$), a maximum prediction error variance $\sigma_{\max}^2$ occurs. By using the assumption $p_{00}^{T_{up}} \gg 1/2$, equation (10) can be simplified as follows:

$$\begin{aligned}&c(S_{t_i}, T\,|\,\hat{\theta}_{S_{t_i}} = 0)\\ &\approx (1-P_e(S_{t_i}))\sum_{\ell=1}^{T-1}(1-\alpha^{2\ell})\\ &\quad +P_e(S_{t_i})(T-1)\sigma_{\max}^2 + \frac{\rho}{T}. \qquad (11)\end{aligned}$$

The sampling state cost function conditioned on $\hat{\theta}_{S_{t_i}} = 1$ (i.e., $c(S_{t_i}, T\,|\,\hat{\theta}_{S_{t_i}} = 1)$) can be written similarly.

Finally, for the case $\hat{\theta}_{S_{t_i}} = 2$, we assume that the prediction variance is the maximum prediction error variance $\sigma_{\max}^2$:

$$c(S_{t_i}, T\,|\,\hat{\theta}_{S_{t_i}} = 2) = (T-1)\sigma_{\max}^2 + \frac{\rho}{T}. \qquad (12)$$

We analyze the performance of the proposed greedy sampling scheme in the following theorems. We derive upper and lower bounds for average sampling rate and expected reconstruction distortion per sample of the proposed sampling scheme. Simulation results for this sampling scheme are presented in Section VI, which support the derived analytical bounds.

Before presenting theorems, we introduce some notations. Suppose that, for all S_{t_i}, $P_e^{\text{low}} \leq P_e(S_{t_i}) \leq P_e^{\text{up}}$. By considering an upper bound on $P_e(S_{t_i})$, we define

$$\begin{aligned}T_0^{\text{low}} = \arg\min_T \quad &(1-P_e^{\text{up}})\sum_{l=1}^{T-1}(1-\alpha_0^{2l})\\ &+P_e^{\text{up}}(T-1)\sigma_{\max}^2 + \frac{\rho}{T}. \qquad (13)\end{aligned}$$

T_0^{up} is defined similarly by considering a lower bound on $P_e(S_{t_i})$. Analogously, T_1^{up} and T_1^{low} can be defined.

Also, d_0^{up}, an upper bound on the expected reconstruction distortion per sample given $\hat{\theta}_{S_{t_i}} = 0$ is defined as follows:

$$d_0^{\text{up}} = \frac{1}{T_0^{\text{up}}}\big\{(1-P_e^{\text{up}})\sum_{\ell=1}^{T_0^{up}-1}(1-\alpha_0^{2\ell})+P_e^{up}(T_0^{up}-1)\sigma_{\max}^2\big\}.$$

Quantities d_0^{low}, d_1^{up} and d_1^{low} are defined similarly.

The following theorem provides analytical upper and lower bounds on the average sampling rate and the expected reconstruction distortion of the greedy sampling scheme introduced in Algorithm 6.

Theorem 8: Consider a Markovian signal defined in (3) over a large enough time interval $[0, T_{\text{tot}}]$. Under Assumption 7, an achievable rate-distortion pair (R, D) of the greedy sampling scheme of Algorithm 6 can be bounded as follows:

$$\begin{aligned}\frac{1}{2T_0^{\text{up}}}+\frac{1}{2T_1^{\text{up}}} \leq \quad R \quad &\leq \frac{1}{2T_0^{\text{low}}}+\frac{1}{2T_1^{\text{low}}} \qquad (14)\\ \frac{d_0^{\text{low}}}{2}+\frac{d_1^{\text{low}}}{2} \leq \quad D \quad &\leq \frac{d_0^{\text{up}}}{2}+\frac{d_1^{\text{up}}}{2}. \qquad (15)\end{aligned}$$

Proof: See Section VII-E. ■

Similarly to Corollary 5, T_0^{low}, T_0^{up}, T_1^{low} and T_1^{up} can be calculated by finding roots of some equations. For example:

Corollary 9: Suppose T_{root} is a solution of the following equation:

$$(1-P_e^{\text{up}})(1-\alpha_0^{2T})+P_e^{\text{up}}(T-1)\sigma_{\max}^2-\frac{\rho}{T(T+1)} = 0 \quad (16)$$

where $0 < \alpha < 1$ and $\rho > 0$. If $(1-P_e^{\text{up}})(1-\alpha_0^2) < \frac{\rho}{2}$, then,

(i) T_{root} is unique; and
(ii) $T_0^{\text{low}} = \lfloor T_{\text{root}} \rfloor$ or $T_0^{\text{low}} = \lfloor T_{\text{root}} \rfloor + 1$.

A similar corollary can be stated for T_0^{up}, T_1^{low} and T_1^{up}.

In Theorem 8, the performance of the proposed sampling scheme (i.e., its average sampling rate and the expected reconstruction distortion) is bounded. However, it is insightful to compare its performance to a genie-aided sampling scheme where the state of the underlying Markov chain is known. For a genie-aided scheme, $P_e^{\text{low}} = P_e^{\text{up}} = 0$, and therefore upper and lower bounds of Theorem 8 match:

Corollary 10: For a genie-aided sampling scheme, the following rate-distortion pair is achievable:

$$\begin{aligned}R &= \frac{1}{2T_0^{\text{genie}}}+\frac{1}{2T_1^{\text{genie}}}, \qquad (17)\\ D &= \frac{d_0^{\text{genie}}}{2}+\frac{d_1^{\text{genie}}}{2}, \qquad (18)\end{aligned}$$

where

$$T_0^{\mathrm{genie}} = \arg\min_T \sum_{\ell=1}^{T-1}(1-\alpha_0^{2\ell}) + \frac{\rho}{T},$$
$$d_0^{\mathrm{genie}} = \frac{1}{T_0^{\mathrm{genie}}}\Big\{\sum_{\ell=1}^{T_0^{\mathrm{genie}}-1}(1-\alpha_0^{2\ell})\Big\},$$

and T_1^{genie} and d_1^{genie} are defined similarly.

The proof of Corollary 9 is similar to the one of Corollary 5. Also, Corollary 10 can be derived by using $P_e^{\mathrm{low}} = P_e^{\mathrm{up}} = 0$ in Theorem 8.

In the first step of this sampling function, we need to estimate $\theta_{S_{t_i}}$ by using m most recently taken samples of the signal and compute its probability of error $P_e(\hat{\theta}_{S_{t_i}})$. In the following, we give an example of such a scheme.

Example 11: Suppose $m = 2$ and we use a maximum likelihood estimator. If $\theta_{S_{t_i}} = 0$, the probability distribution of $X(t_i)$ is $\mathcal{N}(\alpha_0^{T_{i-1}}X(t_{i-1}), 1-\alpha_0^{2T_{i-1}})$, where $\mathcal{N}(\mu,\sigma^2)$ represents a Gaussian distribution with mean μ and variance σ^2. The prior probability of this event is $p_{00}^{T_{i-1}}$. Similarly, if $\theta_{S_{t_i}} = 1$, $X(t_i)$ is distributed by $\mathcal{N}\big(\alpha_1^{T_{i-1}}X(t_{i-1}), 1-\alpha_1^{2T_{i-1}}\big)$. The prior probability of this event is $p_{11}^{T_{i-1}}$. Otherwise, the distribution of $X(t_i)$ is $\mathcal{N}(\alpha_0^j\alpha_1^{T_{i-1}-j}X(t_{i-1}), 1-\alpha_0^{2j}\alpha_1^{2(T_{i-1}-j)})$ with the prior probability $\binom{T_{i-1}}{j}p_{00}^j p_{11}^{T_{i-1}-j}$, for $1 \leq j \leq T_{i-1}-1$. Each of these events corresponds to the case $\theta_{S_{t_i}} = 2$ (i.e., there is a transition within the sampling state.). Therefore, by having the observed value of $X(t_i)$ and using these distributions, a maximum likelihood estimator can estimate $\hat{\theta}_{S_{t_i}}$ and compute its error probability.

Remarks:

1) If the state space is not large, computations can be performed off-line and results can be used in the sampling function.
2) When the order of the sampling function is large and/or autocorrelation coefficients change continuously, a maximum likelihood estimator may not be practically interesting. In these cases, we can use previously taken samples within a window of size W from the last sample (i.e., all taken samples from time $t_i - W + 1$ to the time t_i) to update or estimate autocorrelation coefficients to use in the sampling function. The quality of this estimation process depends on the window size W, the variation rate of autocorrelation coefficients, and the technique used. Two possible methods for estimating autocorrelation coefficients are as follows:
   - A gradient-based method. Suppose at the sampling state $S_{t_{i-1}}$, the set of estimated autocorrelation coefficients is $\{\hat{r}(j) : j \geq 1\}$. By taking a sample at time n_i, these coefficients are updated as follows:
     $$\hat{r}(j) := \hat{r}(j) + \gamma(X(t_i)X(t_i - j) - \hat{r}(j)) \qquad (19)$$
     for all possible j's, where $:=$ represents an update sign, and $\gamma > 0$ is a gradient step size. This gradient-based update method can be useful when W is not large.
   - A window-based method. If the window size W is large and there are enough known samples within the window, an empirical value for each autocorrelation coefficient can be computed.

The estimated autocorrelation coefficients $\{\hat{r}(j) : j \geq 0\}$ can be used in the generalized linear prediction filter in order to design a sampling function similar to the one of Algorithm 6.

V. Dynamic Programming-Based TANS

In greedy TANS, sampling functions are derived based on minimizing the sampling cost at each sampling state. Hence, it does not take into account the *quality* of next sampling states. Intuitively, the larger the sampling increment at the sampling state S_{t_i}, the lower the quality of the next sampling state. Therefore, in general, greedy methods may not provide optimal sampling functions with respect to the overall sampling cost.

We consider *quality* of next sampling states in dynamic programming-based TANS methods. For the sampling state S_{t_i}, a cost-to-go function $J_f(S_{t_i})$ is defined as follows:

$$\begin{aligned} J_f(S_{t_i}) &= c(S_{t_i}, T_i) + \beta c(S_{t_{i+1}}, T_{i+1}) + \cdots \\ &= \sum_{j \geq i} \beta^{j-i} c(S_{t_j}, T_j). \end{aligned} \qquad (20)$$

In this setup, $0 < \beta < 1$ is called a *discount factor*. An interpretation of this factor is that the cost of the next sampling state is less important for the current state policy by a factor of β. Another interpretation of this factor is that the process may be ended at each sampling state with probability $1-\beta$. An optimal cost-to-go function $J(S_{t_i})$ of the sampling state S_{t_i} is defined as follows:

$$J(S_{t_i}) = \inf_f J_f(S_{t_i}). \qquad (21)$$

A Bellman-Ford equation (BFE) can be written for this problem by using cost-to-go functions of different sampling states as follows:

$$J(S_{t_i}) = \inf_{T_i} \big(c(S_{t_i}, T_i) + \beta E[J(S_{t_{i+1}})]\big). \qquad (22)$$

A solution of this Bellman-Ford equation (BFE) is an optimal solution for the sampling problem presented in Definition 1 when the reconstruction function is causal. We investigate this problem for various signal models and sampling setups in this section. For some cases where the number of sampling states is not large, an optimal sampling function can be derived. In other cases where finding this optimal solution is computationally difficult, we propose sampling functions based on approximate dynamic programming (ADP) algorithms. We define a *quality* function $q(S_{t_i})$ for each sampling state S_{t_i}. A greedy solution is used to define this quality function. Then, a sampling function can be computed as follows:

$$\begin{aligned} T_i = \arg\inf_T \quad & c(S_{t_i}, T) + \beta E[q(S_{t_{i+1}})] \\ \text{subject to} \quad & T \geq 1. \end{aligned} \qquad (23)$$

In this setup, we consider quality effects of just one sampling state ahead. Also, note that the solution of the BFE is optimal if the reconstruction function $g(\cdot)$ is causal. It is a necessary assumption of the dynamic programming setup to have a separation of sampling state costs at different stages.

A. An Online Source Coding Scheme Based on TANS

In this section, we consider a sampling problem where an exact DP-based solution can be derived as a solution of the BFE (22). Consider a Markovian signal with an underlying hidden Markov chain depicted in Figure 3. For simplicity, suppose $p_{01} = \epsilon_0 \ll 1$ and $p_{10} = \epsilon_1 \ll 1$. Other regimes of transition probabilities can be analyzed similarly. Suppose $X(t)$ is a binary signal generated by this underlying hidden Markov chain so that, at state 0, $X(t) = 0$, and at state 1, $X(t) = 1$. For the reconstruction, we use a causal function that selects a most-probable binary sequence to fill missing places. Hamming distance is used as an error measure. We call this problem an *online source coding* problem since it does not have a compression delay as in an unconstrained block source coding scheme. In this problem, TANS can provide some compression gain without having any delay. This problem can be extended to Markov chains with more states and different regimes of transition probabilities.

In this sampling problem, there are two different sampling states: If $X(t_i) = 0$, then, $S_{t_i} = 0$; otherwise, $S_{t_i} = 1$. Suppose $f(\cdot)$ is the sampling function (i.e., $T_i = f(S_{t_i})$.). Suppose $f(S_{t_i} = 0) = T_0$ and $f(S_{t_i} = 1) = T_1$. Hence, T_0 and T_1 represent sampling steps at different sampling states.

Since $\epsilon_0, \epsilon_1 \ll 1$, the reconstruction method would choose all-0 and all-1 sequences when $S_{t_i} = 0$ and $S_{t_i} = 1$, respectively. Suppose $S_{t_i} = 0$. Hence, the next sample is taken at time $t_i + T_0$. An error happens if there is one or more Markov chain transitions over the time interval $[t_i + 1, t_i + T_0 - 1]$. To simplify our analysis, we only consider first-order error terms (i.e., we assume at most one transition happens over a time interval $[t_i + 1, t_i + T_0 - 1]$). To have this simplifying assumption, we need to restrict the sampling increments such that $\max T_0, T_1 \ll \min\{\frac{1}{\epsilon_0}, \frac{1}{\epsilon_1}\}$. Therefore, the probability of having more than one Markov chain transition over a time interval of a length T_0 or T_1 is negligible.

By considering Hamming distance as a distortion measure, the sampling state cost at $S_{t_i} = 0$ can be written as follows:

$$c(S_{t_i}, T_0) = \sum_{j=1}^{T_0-1} (1-\epsilon_0)^{j-1}\epsilon_0(T_0 - j) + \frac{\rho}{T_0}. \tag{24}$$

Note that, $(1-\epsilon_0)^{j-1}\epsilon_0$ is the probability of not having a transition over the first $j-1$ samples of the sampling interval and having a transition at the jth sample. Hence, the considered reconstruction method makes errors on samples from time t_i+j to t_i+T_0-1, which corresponds to a Hamming distance of $T_0 - j$. A similar argument can be made for the case of $S_{t_i} = 1$.

By considering these sampling state cost functions, the BFE can be written as follows:

$$\begin{aligned} J(S_{t_i} = 0) &= \min_{T_0,T_1} \big[c(S_{t_i} = 0, T_0) \\ &\quad +\beta(1-\epsilon_0)^{T_0} J(S_{t_i} = 0) \\ &\quad +\beta(1-(1-\epsilon_0)^{T_0}) J(S_{t_i} = 1)\big], \\ J(S_{t_i} = 1) &= \min_{T_0,T_1} \big[c(S_{t_i} = 1, T_1) \\ &\quad +\beta(1-\epsilon_1)^{T_1} J(S_{t_i} = 1) \\ &\quad +\beta(1-(1-\epsilon_1)^{T_1}) J(S_{t_i} = 0)\big]. \end{aligned}$$

Since these BFEs have only two variables, various numerical and analytical methods can be applied to find their solution, which in turn corresponds to an optimal sampling scheme in TANS (e.g., see [17]). Simulation results for this sampling scheme are given in Section VI.

B. Approximate Dynamic Programming Methods

Finding a solution of the BFE of a DP-based sampling function may not be practically feasible if the sampling state space is large or is unknown (or is known partially). In these cases, approximate dynamic programming (ADP) algorithms can provide suboptimal solutions with a reasonable computation complexity. In this section, we investigate an ADP-based sampling function of TANS introduced in (23). To use this approximate algorithm, for any sampling state S_{t_i}, a *quality* measure $q(S_{t_i})$ is assigned. We use a greedy sampling solution to define this quality function.

Consider a Markovian signal $X(t)$ described in Section III-B. A greedy sampling function for this signal is introduced in Algorithm 6. We refer to this greedy sampling function as $f^{\text{greedy}}(\cdot)$, where $T_i^{\text{greedy}} = f^{\text{greedy}}(S_{t_i})$.

A *quality* function of each state is defined as

$$q(S_{t_i}) = \gamma T_i^{\text{greedy}} \tag{25}$$

where γ is a scaling parameter. Intuitively, the larger the greedy sampling step, the higher the quality of the sampling state. Therefore, an ADP-based sampling function of TANS can be derived by using the optimization setup of (23).

For computing the expected quality of the next sampling state, transition probabilities among different sampling states should be known. Suppose, at a sampling state S_{t_i}, the next sample is taken after a time interval T_i. The value of the sample $X(t_i+T_i)$ is a random variable with a mean $\widehat{X}(t_i+T_i)$, which can be computed by a generalized linear prediction filter. The probability distribution of $X(t_i + T_i)$ determines the probability distribution of the next sampling state $S_{t_{i+1}}$. Therefore, the expected value of the quality function of the next state can be computed by using this probability distribution. However, to simplify this sampling function further, one may approximate this expected quality by the quality of the most probable next state, which has a sample value of $\widehat{X}(t_i + T_i)$ at time t_{i+1}. We call this state $\widehat{S}_{t_{i+1}}$. Therefore, a more simplified sampling function based on an ADP algorithm can be written as

$$T_i \quad = \quad \arg\min_{T} \quad c(S_{t_i}, T) + \beta q(\widehat{S}_{t_{i+1}}).$$

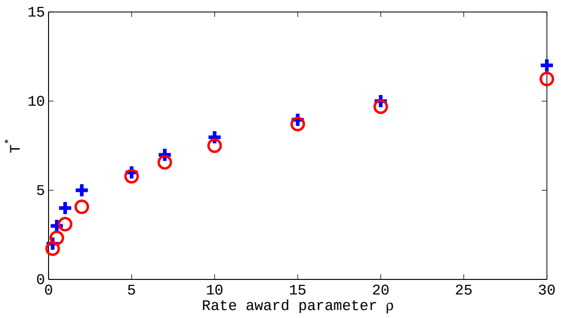

Fig. 4. Illustration of Corollary 5 by simulation. The red curve is the solution of (9). The blue curve is optimal sampling rate T^*. Note that, their difference is always less than 1.

For any given state S_{t_i} and $T_i = T$, $q(\widehat{S}_{t_{i+1}})$ can be computed. Intuitively, the term $\beta q(\widehat{S}_{t_{i+1}})$ is a correction term for the greedy solution considering the quality of the next sampling state. In this scheme, the quality of only one future sampling state is considered. However, one can extend this algorithm to consider the quality of more than one future sampling state.

Algorithm 12: An approximate dynamic programming-based sampling function for the Markovian signal of (3) can be summarized as follows:

- Step i,0: Compute $\hat{\theta}_{S_{t_i}}$, $P_e(S_{t_i})$ and $q(\widehat{S}_{t_{i+1}})$.
- Step i,1: Compute $T_i = \min c(S_{t_i}, T|\hat{\theta}_{S_{t_i}}) + \beta q(\widehat{S}_{t_{i+1}})$.
- Step i,2: Take a sample at time $t_i + T_i$.
- Step i,3: Compute S_{t_i+1}. Repeat.

Simulation results for this sampling procedure are presented in Section VI.

VI. Simulation Results

In this section, we evaluate the performance of the proposed sampling schemes by simulations and compare their performance against uniform sampling. In uniform sampling, the sampling rate is always in the form of $R = 1/T_{\text{uni}}$, , where T_{uni} is a positive integer. To be able to compare the performance of different methods with uniform sampling at different rates, we modify the uniform sampling setup to capture all possible sampling rates. To do this, for a given rate $R = 1/T_{\text{uni}}$ where T_{uni} is not an integer number, the ith sample is taken at time $t_i = \text{round}(T_{\text{uni}})$.

First, we consider an autoregressive signal model introduced in (2). In Theorem 4, we show that an optimal sampling scheme for this signal is uniform. Also, Corollary 5 provides a straightforward way to compute the optimal sampling rate. Figure 4 illustrates this corollary for the case of $\alpha = 0.99$ (i.e., noise power is 0.02) and for different ρ values. As shown in this figure, the difference between the root of equation (9) and the optimal increment T^* is always less than 1.

Now, we consider Markovian signals introduced in (3). Transition probabilities of the underlying MC are assumed to be 0.001. We demonstrate the performance of different sampling methods on a rate-distortion plots (Figures 5, 6 and 7). Rate refers to average sampling rate and distortion refers to average reconstruction distortion per sample. Lower curves in these plots indicate better performance.

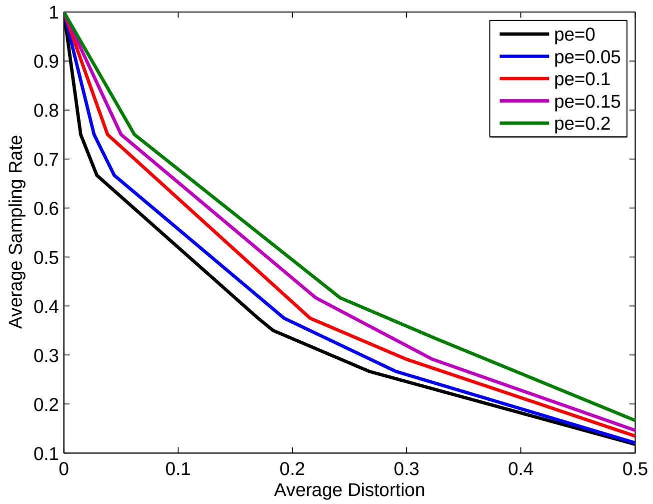

Fig. 5. Analytical rate-distortion curves of Theorem 8 for a Markovian signal of equation (3) with the signal parameters $\alpha_0 = 0.01$ and $\alpha_1 = 0.99$. P_e represents error probability of the estimation process. $P_e = 0$ curve corresponds to the so-called genie-aided scheme.

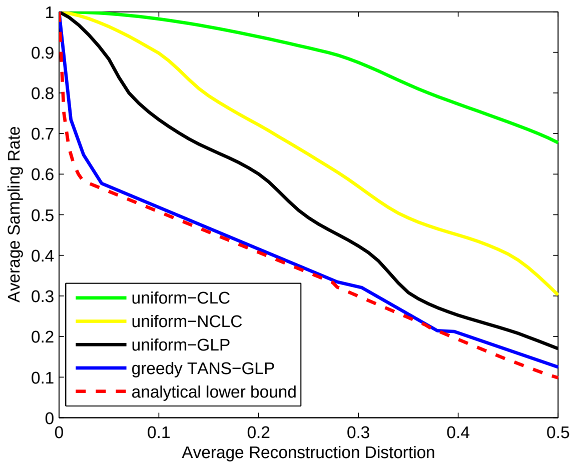

Fig. 6. Average sampling rate versus average reconstruction distortion for a Markovian signal with parameters $\alpha_0 = 0.01$ and $\alpha_1 = 0.99$ for methods: (i) uniform sampling with causal line-connecting (CLC) reconstruction, (ii) uniform sampling with non-causal line-connecting (NCLC) reconstruction, (iii) uniform sampling with generalized linear prediction (GLP) filtering (iv) greedy TANS with generalized linear prediction (GLP) filtering, and (v) analytical lower bound for greedy TANS based on Theorem 8.

Figure 5 shows analytical rate-distortion curves of Theorem 8 for a Markovian signal of equation (3) for different estimation error probabilities. Here, noise power in state 0 of the Markov chain is 0.05 (i.e., $\alpha_0 \approx 0.97$) and the noise power in state 1 is 0.5 (i.e., $\alpha_1 \approx 0.7$). As illustrated in this plot, the lower the error probability, the better the performance. Note that the case where $P_e = 0$ is referred as a genie-aided scheme.

Figures 6 and 7 show rate-distortion curves achieved by simulations for various schemes. The signal model is Markovian (3) with parameters $\alpha_0 = 0.01$ and $\alpha_1 = 0.99$ in schemes considered in Figure 6 and, $\alpha_0 = 0.7$ and $\alpha_1 = 0.97$ in the ones of Figure 7. In greedy schemes, we use a maximum

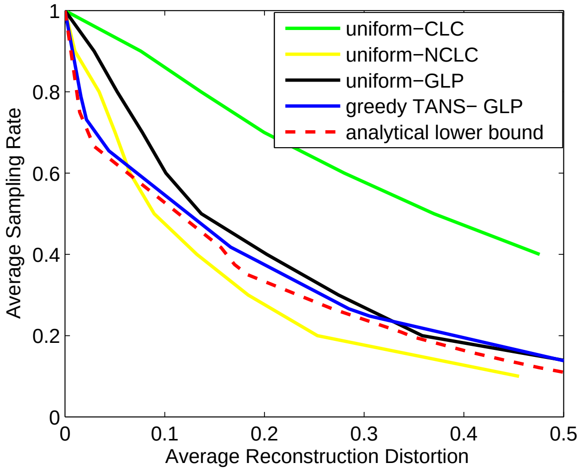

Fig. 7. Average sampling rate versus average reconstruction distortion for a Markovian signal with parameters $\alpha_0 = 0.7$ and $\alpha_1 = 0.97$ for methods: (i) uniform sampling with causal line-connecting (CLC) reconstruction, (ii) uniform sampling with non-causal line-connecting (NCLC) reconstruction, (iii) uniform sampling with generalized linear prediction (GLP) filtering (iv) greedy TANS with generalized linear prediction (GLP) filtering, and (v) analytical lower bound for greedy TANS based on Theorem 8.

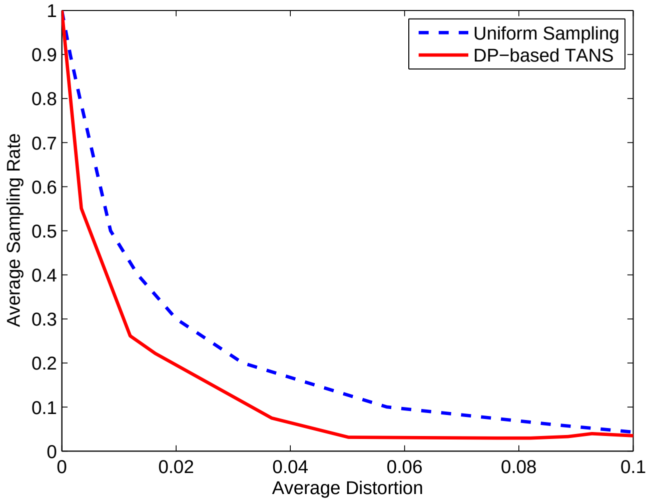

Fig. 8. A rate-distortion plot of a dynamic programming-based TANS for an online source coding application explained in Section V-A.

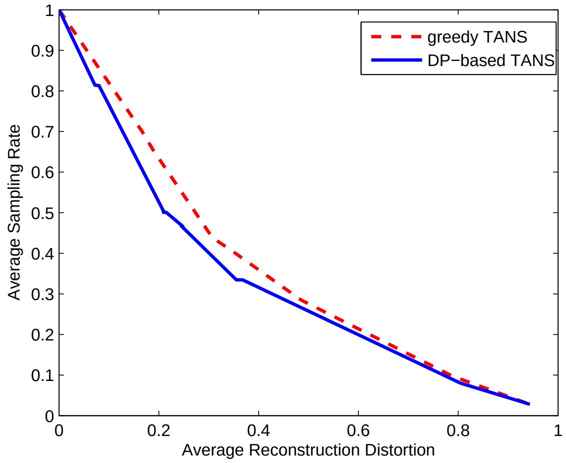

Fig. 9. Comparison of a dynamic programming-based TANS with greedy TANS for a Markovian signal model.

likelihood estimation block with $m = 10$ to estimate the state of the underlying Markov chain. For greedy TANS, we use generalized linear prediction (GLP) filter as the reconstruction method. Note that, this reconstruction is causal. For uniform sampling, we use three reconstruction methods: causal line-connecting (CLC), non-causal line-connecting (NCLC) and GLP filtering. Although, it is not fair to compare performance of greedy TANS with causal reconstruction with a uniform sampling scheme with non-causal reconstruction, in the case of Figure 6, greedy TANS outperforms uniform sampling schemes including the one with a non-causal reconstruction method. In the case of Figure 7, greedy TANS outperforms uniform schemes with causal reconstructions. In the low-distortion regime (distortion less than 0.08), it also outperforms uniform sampling with non-causal reconstruction.

As illustrated in these figures, genie-aided greedy TANS provides an analytical lower bound for greedy TANS in the rate-distortion plot. The proposed greedy TANS performs closely to this lower bound. Also, by choosing the estimation error probability 0.05 (estimated from simulations), an analytical upper bound for greedy TANS can be achieved as proposed in Theorem 8. Moreover, by comparing two greedy TANS schemes of these figures, we notice that the more different α values of MC states, the more gain is provided by TANS framework.

Figure 8 shows the performance of a dynamic programming-based TANS scheme for an online source coding application explained in Section V-A. Here, we assume $\epsilon_0 = 0.1$ and $\epsilon_1 = 0.01$. To solve the Bellman-Ford equation (25), a value-iteration method is used [17]. As shown in this plot, a DP-based TANS scheme outperforms uniform sampling.

Figure 9 illustrates performance of a TANS scheme based on approximate dynamic programming for a Markovian signal explained in Algorithm 12. Here, we assume that underlying Markov chain transition probabilities are 0.1 (i.e., $p_{01} = p_{10} = 0.1$). The signal parameters are assumed to be $\alpha_0 = 0.7$ and $\alpha_1 = 0.99$. As illustrated in this figure, a TANS scheme based on ADP outperforms the greedy one.

VII. Proofs

A. Proof of Theorem 2

To find a solution of the optimization problem (4), we use similar techniques as for the regular linear prediction filter [16]. Note that in an optimal scheme the error term should be orthogonal to all known samples:

$$E\big[X(t_{i-k})e^*(t)\big] = 0 \tag{26}$$

for $k = 0, \ldots, m-1$, where $e^*(t) = X(t) - \sum_{k=0}^{m-1} w^*_{\tau_k} X(t_{i-k})$.

By using (26) and (5), optimal weights $w^*_{\tau_i}$, for $0 \leq i \leq$

$m-1$ should satisfy the following set of linear equations:

$$r(-\tau_k)=\sum_{i=0}^{m-1} w^*_{\tau_i} r(\tau_i-\tau_k) \tag{27}$$

for $k=0,\ldots,m-1$. By using matrix notations of (6), linear equations of (27) can be written as follows:

$$\mathbf{p}=\mathbf{R}\mathbf{w}^*_{\tau}.$$

B. Proof of Theorem 3

Since $X(t)$ has zero mean and $X(t)=\widehat{X}(t)+e^*(t)$, by using (26), we have

$$\sigma^2_{e^*}(S_{t_i},T)=\sigma^2_X-\sigma^2_{\widehat{X}}=r(0)-\sigma^2_{\widehat{X}}, \tag{28}$$

where $\widehat{X}(t)=(\mathbf{w}^*_{\tau})^T(\mathbf{X}_{S_{t_i}})$. Therefore,

$$\begin{aligned}\sigma^2_{\widehat{X}} &= E[|\widehat{X}(t)|^2]\\ &= (\mathbf{w}^*_{\tau})^T E\big[(\mathbf{X}_{S_{t_i}})(\mathbf{X}_{S_{t_i}})^T\big]\mathbf{w}^*_{\tau}\\ &= (\mathbf{w}^*_{\tau})^T\mathbf{R}\mathbf{w}^*_{\tau}\\ &= \mathbf{p}^T\mathbf{w}^*_{\tau}.\end{aligned} \tag{29}$$

Equations (28) and (29) establish the theorem.

C. Proof of Theorem 4

By using the definition of an AR(1) signal, the sampling state cost can be written as follows:

$$c(S_{t_i},T)=\sum_{j=1}^{T-1}(1-\alpha^{2j})+\frac{\rho}{T}. \tag{30}$$

Hence, an optimal greedy sampling solution for this sampling state cost can be computed as follows:

$$T^*_{\text{greedy}}=\arg\min_T \sum_{j=1}^{T-1}(1-\alpha^{2j})+\frac{\rho}{T}. \tag{31}$$

Now, we show that this optimal greedy TANS (which is a uniform sampling scheme) performs closely to an optimal sampling scheme, which may be nonuniform. Consider a uniform sampling state S_{t_i} with an inter-state sampling increment T^*_{greedy}. The sampling state cost of this uniform sampling scheme is referred to by $c_{\min}$ (this sampling state cost can be achieved by having T^*_{greedy} as a solution of the optimization setup (7)). First, we show that this uniform sampling scheme satisfies the BFE (22). Since under this sampling scheme, $S_{t_{i+1}}=S_{t_i}$, therefore, the cost-to-go function at the sampling state S_{t_i} can be written as follows:

$$\begin{aligned}J_f(S_{t_i}) &= c(S_{t_i},T_i)+\beta c(S_{t_{i+1}},T_{i+1})+\cdots\\ &= c(S_{t_i},T^*_{\text{greedy}})\big(1+\beta+\beta^2+\cdots\big)\\ &= \frac{c_{\min}}{1-\beta}.\end{aligned} \tag{32}$$

Now, consider the right hand side (RHS) of the BFE (22):

$$\begin{aligned}\text{RHS of BFE} &= \min_{T_i}\big(c(S_{t_i},T_i)+\beta E[J_f(S_{t_{i+1}})]\big)\\ &= c_{\min}+\beta\frac{c_{\min}}{1-\beta}\\ &= \frac{c_{\min}}{1-\beta}\\ &= J_f(S_{t_i}).\end{aligned} \tag{33}$$

Therefore, the above uniform sampling scheme satisfies the BFE when S_{t_i} (i.e., the initialization state) happens to be a uniform sampling state with an inter-state sampling step size T^*_{greedy}. If the process starts from another sampling state, we assume that there is always a way to reach to this uniform sampling state (e.g., take m uniform samples with a sampling increment T^*_{greedy}). Moreover, the mapping functions between policies and costs are continuous. Hence, a small difference in costs due to initialization effects has a small effect in sampling policies. Therefore, for an AR(1) signal model, the optimal greedy TANS (which is uniform) performs closely to an optimal sampling scheme (which may be nonuniform).

D. Proof of Corollary 5

Define $h(T)=(1-\alpha^{2T})-\rho/(T(T+1))$ for $T\in[1,\infty)$. Note that $h(T)$ is a continuous function over this interval. Since $h(1)=(1-\alpha^2)-\rho/2<0$ and $\lim_{T\to\infty}h(T)=1>0$, by using the mean value theorem, h has a root T_{root} in $[1,\infty)$. Since $dh(T)/dT>0$, this root is unique, completing the proof of part (i).

For an AR(1) signal, we have:

$$h(T)=c(S_{t_i},T+1)-c(S_{t_i},T)=(1-\alpha^{2T})-\frac{\rho}{T(T+1)}.$$

Since $T^*=\arg\min_T c(S_{t_i},T)$ and also T^* is an integer, either $T^*=\lfloor T_{\text{root}}\rfloor$ or $T^*=\lfloor T_{\text{root}}\rfloor+1$, completing the proof of part (ii).

E. Proof of Theorem 8

Under conditions of (10), (11) shows the sampling state cost given $\hat{\theta}_{S_{t_i}}=0$. By using the sampling scheme proposed in Algorithm 6, we have:

$$T_i=\arg\min_T c(S_{t_i},T|\hat{\theta}_{S_{t_i}}=0). \tag{34}$$

Since $\sigma^2_{\max}\geq 1$, by having $P_e^{\text{low}}\leq P(S_{t_i})\leq P_e^{\text{up}}$, given $\hat{\theta}_{S_{t_i}}=0$, we have: $T_0^{\text{low}}\leq T_i\leq T_0^{\text{up}}$. Similarly, one can show that, given $\hat{\theta}_{S_{t_i}}=1$, $T_1^{\text{low}}\leq T_i\leq T_1^{\text{up}}$. Since the underlying Markov chain is symmetric (i.e., $p_{01}=p_{10}$), in steady state, $\hat{\theta}_{S_{t_i}}=0$ approximately half of the time and $\hat{\theta}_{S_{t_i}}=1$ approximately half of the time. Hence, the number of samples taken in that state by using the proposed sampling scheme of Algorithm 6 is bounded between

$$\frac{T_{\text{tot}}}{2T_0^{\text{low}}}+\frac{T_{\text{tot}}}{2T_1^{\text{low}}} \quad\text{and}\quad \frac{T_{\text{tot}}}{2T_0^{\text{up}}}+\frac{T_{\text{tot}}}{2T_1^{\text{up}}}.$$

This demonstrates the sampling rate bounds. By using bounds on T_i and $P_e(S_{t_i})$ in (11), deriving bounds on the expected reconstruction distortion is straightforward.

VIII. CONCLUSION

In this paper, we introduced a new framework for an adaptive nonuniform sampling scheme called *time-stampless adaptive nonuniform sampling* (TANS). The key idea of this framework is that, *time increments between samples are computed by using a function of previously taken samples*. Therefore, keeping sampling times (time stamps), except initialization times, is not necessary. We introduced two methods to design sampling functions for discrete-time stochastic signals: a greedy method, and a method based on dynamic programming. We analyzed the performances of these schemes by computing (or bounding) their trade-offs between sampling rate and expected reconstruction distortion for autoregressive and Markovian signals. We showed that, by being time-stampless and opportunistically adapting to local signal characteristics, TANS can provide significant rate-distortion gains, which can be translated to improved power efficiency in some applications.

IX. ACKNOWLEDGMENT

Authors would like to thank Muriel Rambeloarison for helping in some of the simulations.

REFERENCES

[1] R. H. Walden, “Analog-to-digital converter survey and analysis,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, 1999.

[2] F. A. Marvasti, *Nonuniform Sampling: Theory and Practice*. Plenum Publishers Co., 2001.

[3] S. M. Qaisar, L. Fesquet, and M. Renaudin, “Computationally efficient adaptive rate sampling and filtering,” in *15th European Signal Processing Conference (EUSIPCO 07)*, Poznan, Poland, Sep. 2007, pp. 2139–2143.

[4] D. Wei, “Sampling based on local bandwidth,” in *Master Thesis, MIT*, 2007.

[5] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, “On stationary linear time-varying systems,” *IEEE Trans. Circ. Syst.*, vol. 29, no. 3, pp. 169–184, Mar. 1982.

[6] K. Horiuchi, “Sampling principle for continuous signals with time-varying bands,” *Inform. and Control*, vol. 13, no. 1, pp. 53–61, 1968.

[7] M. Unser and J. Zerubia, “A generalized sampling theory without bandlimiting constraints,” *IEEE Trans. Circ. Syst.*, vol. 2, pp. 959–969, 1998.

[8] P. P. Vaidyanathan and B. Vrcelj, “On sampling theorems for non bandlimited signals,” in *Proc. of the ICASSP*, Salt Lake City, May 2001, pp. 3897–3900.

[9] J. D. Gibson, “Sequentially adaptive backward prediction in ADPCM speech coders,” *IEEE Trans. Commun.*, vol. 26, no. 1, pp. 145–150, Jan. 1978.

[10] N. S. Jayant, “Adaptive quantization with a one-word memory,” *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1119–1144, Sep. 1973.

[11] J. D. Gibson, “Adaptive prediction in speech differential encoding systems,” *Proc. IEEE*, vol. 68, no. 4, pp. 488–525, Apr. 1980.

[12] A. Ortega and M. Vetterli, “Adaptive scalar quantization without side information,” *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 665–676, May 1997.

[13] V. K. Goyal, J. Zhuang, and M. Vetterli, “Transform coding with backward adaptive updates,” *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1623–1633, Jul. 2000.

[14] T. Lai, “Sequential analysis: some classical problems and new challenges,” *Statistica Sinica*, vol. 11, no. 2, pp. 303–350, 2001.

[15] R. Castro and R. Nowak, “Upper and lower error bounds for active learning,” in *The 44th Annual Allerton Conference on Communication, Control and Computing*, vol. 2, no. 2.1, 2006, p. 1.

[16] S. S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[17] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 1995.